

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 9 (2012) 1234 – 1239

Procedia
 Computer Science

International Conference on Computational Science, ICCS 2012

Regularized Multiple Criteria Linear Programming via Linear Programming

Zhiquan Qi, Yingjie Tian*, Yong Shi*

Research Center on Fictitious Economy & Data Science, Chinese Academy of Sciences, Beijing 100190, China

Abstract

Regularized multiple-criteria linear programming (RMCLP) model is a new powerful method for classification and has been used in various real-life data mining problems. In this paper, instead of solving quadratic programming (QP) in algorithm RMCLP, we proposed a novel RMCLP via linear programming (LP) for pattern classification (called LP-RMCLP). Numerical experiments on benchmark datasets show that the proposed method is comparable to traditional RMCLP in accuracy, however considerably faster than it.

Keywords: Classification, RMCLP, Linear Programming

1. Introduction

Given a independently and identically distributed (i.i.d) training set

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (R^n \times \mathcal{Y})^l, \quad (1)$$

where $x_i \in R^n, y_i \in \mathcal{Y} = \{-1, 1\}, i = 1, \dots, l$. Classification refers to the construction of a discriminate function from the input space R^n onto the unordered set of classes \mathcal{Y} . Recently, Shi and his colleagues [1] proposed a multiple criteria linear programming (MCLP) model for classification. A lot of experiments show the effectiveness of the approach. Then various improved algorithms were proposed one after the other [2, 3, 4, 5, 6] and had been applied to handle many real world data mining problems, such as credit card portfolio management [7, 8, 9], bioinformatics [10], information intrusion and detection [11], firm bankruptcy [12, 13], and etc.

In this paper, we proposed a Regularized Multiple-Criteria Linear Programming Via Linear Programming (called LP-RMCLP). This algorithm can be considered as an improved version of [2]. Their main difference is that our algorithm depends on solving linear programming instead of quadratic programming. Experiment results show that the speed of our algorithm is about 10 times faster than RMCLP on the premise of keeping the classification precision.

We describe our notation now. All vectors will be column vectors unless transposed to a row vector by a prime T . The inner product of two vectors x and y in the n -dimensional real space \mathbb{R}^n will be denoted by $(x \cdot y)$. For $w \in \mathbb{R}^n$, $\|w\|_1$ denotes the 1-norm while $\|w\|_2$ denotes the 2-norm. The notation $B \in \mathbb{R}^{m \times n}$ will signify a real $m \cdot n$

*Corresponding author

Email addresses: qizhiquan@gucas.ac.cn (Zhiquan Qi), tyj@gucas.ac.cn (Yingjie Tian)

matrix. For such a matrix, B^T will denote the transpose of B . A vector of all ones or all zeros of arbitrary dimension will be denoted by e and 0 , respectively. For $x \in \mathbb{R}^n$, the notation $e^T x$ denotes the sum of the components of x .

The remaining parts of the paper are organized as follows. Section 2 introduces the basic formulation of MCLP and RMCLP; In section 3 describes in detail our proposed algorithms LP-RMCLP; All experiment results are shown in the section 4; Last section gives the conclusions.

2. Regularized MCLP for Data Mining

We give a brief introduction of MCLP in the following. For classification about the training data [1]. Data separation can be achieved by two opposite objectives. The first objective separates the observations by minimizing the sum of the deviations (MSD) among the observations. The second maximizes the minimum distances (MMD) of observations from the critical value. the overlapping of data u should be minimized while the distance v has to be maximized. However, it is difficult for traditional linear programming to optimize MMD and MSD simultaneously. According to the concept of Pareto optimality, we can seek the best trade-off of the two measurements[8, 9]. So MCLP model can be described as follows:

$$\min_u \quad e^T u \quad \& \quad \max_v \quad e^T v, \quad (2)$$

$$\text{s.t.} \quad (w \cdot x_i) + (u_i - v_i) = b, \quad \text{for } \{i|y_i = 1\}, \quad (3)$$

$$(w \cdot x_i) - (u_i - v_i) = b, \quad \text{for } \{i|y_i = -1\}, \quad (4)$$

$$u, v \geq 0, \quad (5)$$

where $e \in R^l$ be vector whose all elements are 1, w and b are unrestricted, u_i is the overlapping and v_i the distance from the training sample x_i to the discriminator $(w \cdot x_i) = b$ (classification separating hyperplane). By introducing penalty parameter $c, d > 0$, MCLP has the following version

$$\min_{u,v} \quad ce^T u - de^T v, \quad (6)$$

$$\text{s.t.} \quad (w \cdot x_i) + (u_i - v_i) = b, \quad \text{for } \{i|y_i = 1\}, \quad (7)$$

$$(w \cdot x_i) - (u_i - v_i) = b, \quad \text{for } \{i|y_i = -1\}, \quad (8)$$

$$u, v \geq 0, \quad (9)$$

A lot of empirical studies have shown that MCLP is a powerful tool for classification. However, we cannot ensure this model always has a solution under different kinds of training samples. To ensure the existence of solution, recently, Shi et.al proposed a RMCLP model by adding two regularized items $\frac{1}{2}w^T H w$ and $\frac{1}{2}u^T Q u$ on MCLP as follows(more theoretical explanation of this model can be found in [2]):

$$\min_z \quad \frac{1}{2}w^T H w + \frac{1}{2}u^T Q u + de^T u - ce^T v, \quad (10)$$

$$\text{s.t.} \quad (w \cdot x_i) + (u_i - v_i) = b, \quad \text{for } \{i|y_i = 1\}, \quad (11)$$

$$(w \cdot x_i) - (u_i - v_i) = b, \quad \text{for } \{i|y_i = -1\}, \quad (12)$$

$$u, v \geq 0, \quad (13)$$

where $z = (w^T, u^T, v^T, b)^T \in R^{n+l+l+1}$, $H \in R^{n \times n}$, $Q \in R^{l \times l}$ are symmetric positive definite matrices. Obviously, the regularized MCLP is a convex quadratic programming. The geometric meaning of the model is shown in Figure.1.

3. Regularized Multiple-Criteria Linear Programming Via Linear Programming(LP-RMCLP)

Choose H, Q to be identity matrix, (10)~ (13) can be written as

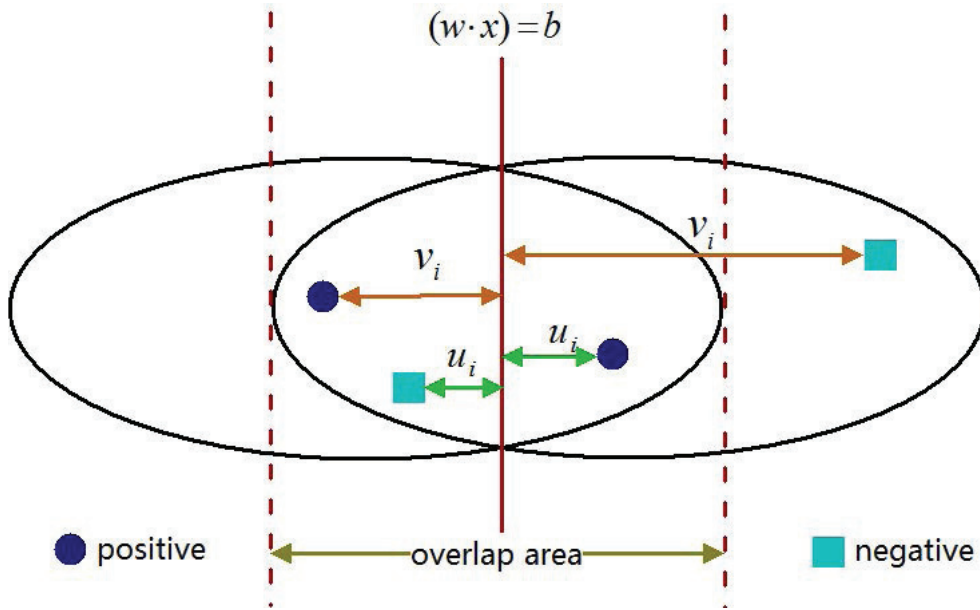


Figure 1: Geometric meaning of MCLP.

$$\min_{w,u,v,b} \quad \frac{1}{2}\|w\|_2^2 + \frac{1}{2}\|u\|_2^2 + d \sum_{i=1}^l u_i - c \sum_{i=1}^l v_i, \quad (14)$$

$$\text{s.t.} \quad y_i((w \cdot x_i) - b) = v_i - u_i, i = 1, \dots, l, \quad (15)$$

$$u_i, v_i \geq 0, i = 1, \dots, l. \quad (16)$$

By introducing its Lagrange function

$$L(w, u, v, b, \alpha, \beta, \eta) = \frac{1}{2}\|w\|_2^2 + \frac{1}{2}\|u\|_2^2 + d \sum_{i=1}^l u_i - c \sum_{i=1}^l v_i + \quad (17)$$

$$\sum_{i=1}^l \alpha_i (y_i((w \cdot x_i) - b) + u_i - v_i) - \sum_{i=1}^l \beta_i u_i - \sum_{i=1}^l \eta_i v_i, \quad (18)$$

where $\alpha_i, \beta_i, \eta_i \in R$ are the Lagrange multipliers, Therefore the dual problem of (14)~ (16) can be formulated as

$$\max_{w,u,v,b,\alpha,\beta,\eta} \quad L(u, v, w, b, \alpha, \beta, \eta), \quad (19)$$

$$\text{s.t.} \quad \nabla_{w,u,v,b} L(u, v, w, b, \alpha, \beta, \eta) = 0, \quad (20)$$

$$\beta_i, \eta_i \geq 0, i = 1, \dots, l. \quad (21)$$

From equation (20) we get

$$\nabla_w L = w + \sum_{i=1}^l y_i \alpha_i x_i = 0, \quad (22)$$

$$\nabla_{u_i} L = u_i + d + \alpha_i - \beta_i = 0, \quad i = 1, \dots, l, \quad (23)$$

$$\nabla_{v_i} L = -c - \alpha_i - \eta_i = 0, \quad i = 1, \dots, l, \quad (24)$$

$$\nabla_b L = - \sum_{i=1}^l y_i \alpha_i = 0. \quad (25)$$

Substituting the above equations into problem (14)~ (16), the dual problem can be expressed as

$$\max_{\alpha, u, b} \quad -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l u_i u_j - \sum_{i=1}^l \alpha_i b, \quad (26)$$

$$\text{s.t.} \quad \sum_{i=1}^l y_i \alpha_i = 0, \quad i = 1, \dots, l, \quad (27)$$

$$-d - u_i \leq \alpha_i \leq -c, \quad i = 1, \dots, l. \quad (28)$$

From (22), we can obtain $w = -\sum_{i=1}^l y_i \alpha_i x_i$. Now we use 1-norm distance $\|\alpha\|_1$ to replace the 2-norm distance $\|w\|_2^2$, and $\|u\|_1$ for $\|u\|_2^2$. So the problem (14)~ (16) can be rewritten as

$$\min_{\alpha, u, v, b} \quad \frac{1}{2} \|\alpha\|_1 + \frac{1}{2} \|u\|_1 + d \sum_{i=1}^l u_i - c \sum_{i=1}^l v_i, \quad (29)$$

$$\text{s.t.} \quad y_i \left(-\sum_{j=1}^l y_j \alpha_j (x_j \cdot x_i) - b \right) = v_i - u_i, \quad i = 1, \dots, l, \quad (30)$$

$$u_i, v_i \geq 0, \quad i = 1, \dots, l. \quad (31)$$

By introducing the constraints $-\lambda \leq \alpha \leq \lambda$ and $-\pi \leq u \leq \pi$, the above problem can be further expressed as

$$\min_{\alpha, u, v, b, \lambda, \pi} \quad \frac{1}{2} \sum_{i=1}^l \lambda_i + \frac{1}{2} \sum_{i=1}^l \pi_i + d \sum_{i=1}^l u_i - c \sum_{i=1}^l v_i, \quad (32)$$

$$\text{s.t.} \quad y_i \left(-\sum_{j=1}^l y_j \alpha_j (x_j \cdot x_i) - b \right) = v_i - u_i, \quad i = 1, \dots, l, \quad (33)$$

$$-\lambda_i \leq \alpha_i \leq \lambda_i, \quad i = 1, \dots, l, \quad (34)$$

$$-\pi_i \leq u_i \leq \pi_i, \quad i = 1, \dots, l. \quad (35)$$

$$u_i, v_i, \lambda_i, \pi_i \geq 0, \quad i = 1, \dots, l. \quad (36)$$

Introduce the kernel function $K(x, x') = (\Phi(x) \cdot \Phi(x'))$ to take place of $(x \cdot x')$, where $\Phi(\cdot)$ is a mapping from the input space R^n to Hilbert space \mathcal{H}

$$\Phi : \begin{array}{l} R^n \rightarrow \mathcal{H}, \\ x \rightarrow \Phi(x). \end{array} \quad (37)$$

Therefore, the linear programming problem (32)~ (36) turns to be

$$\min_{\alpha, u, v, b, \lambda, \pi} \quad \frac{1}{2} \sum_{i=1}^l \lambda_i + \frac{1}{2} \sum_{i=1}^l \pi_i + d \sum_{i=1}^l u_i - c \sum_{i=1}^l v_i, \quad (38)$$

$$\text{s.t.} \quad y_i \left(-\sum_{j=1}^l y_j \alpha_j K(x_j \cdot x_i) - b \right) = v_i - u_i, \quad i = 1, \dots, l, \quad (39)$$

$$-\lambda_i \leq \alpha_i \leq \lambda_i, \quad i = 1, \dots, l, \quad (40)$$

$$-\pi_i \leq u_i \leq \pi_i, \quad i = 1, \dots, l. \quad (41)$$

$$u_i, v_i, \lambda_i, \pi_i \geq 0, \quad i = 1, \dots, l. \quad (42)$$

So we can establish the following algorithm:

Algorithm 1 LP-RMCLP

- 1: Given a mixed training set:
 $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (R^n \times \mathcal{Y})^l$,
 where $x_i \in R^n, y_i \in \mathcal{Y} = \{-1, 1\}, i = 1, \dots, l$;
- 2: Choose appropriate penalty parameters c, d and kernel function K ;
- 3: Construct and solve the problem (38) and (42), then get its solution:
- 4: $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_l)^\top, \hat{u} = (\hat{u}_1, \dots, \hat{u}_l)^\top, \hat{v} = (\hat{v}_1, \dots, \hat{v}_l)^\top$,
 $\hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_l)^\top, \hat{\pi} = (\hat{\pi}_1, \dots, \hat{\pi}_l)^\top, \hat{b}$;
- 5: Obtain the decision function:

$$f(x) = -\sum_{i=1}^l y_i \hat{\alpha}_i K(x_i \cdot x_j) + \hat{b}.$$

4. Numerical Experiments

Our algorithm code was wrote in MATLAB 2010. The experiment environment: Intel Core I5 CPU, 2 GB memory. The “linprog” function with MATLAB is employed to solve linear programming problem related to this paper.

The testing accuracies of all experiments are computed using standard 10-fold cross validation. c, d and RBF kernel parameter σ are all selected from the set $\{2^i | i = -7, \dots, 7\}$ by 10-fold cross validation on the tuning set comprising of random 10% of the training data. Once the parameters are selected, the tuning set was returned to the training set to learn the final decision function.

To demonstrate the capabilities of our algorithm, we report results on UCI dataset. They are respectively “Sonar”, “Ionosphere”, “Australian”, “Pima-Indian”, “CMC”, “Votes”, “WPBC”. In all experiments, out method compared with the standard RMCLP. The table 1 give the experiments results

Table 1: The percentage of tenfold testing accuracy of RMCLP and LP-RMCLP.

Dataset	RMCLP		LP-RMCLP	
	accuracy	time(second)	accuracy	time(second)
Sonar	78.11± 5.42	6.32	79.54± 5.68	45.25
Ionosphere	87.64± 6.83	11.45	89.24± 5.71	149.32
Australian	86.89 ± 5.86	30.12	85.32 ± 5.23	280.46
Pima-Indian	77.71 ± 6.13	34.26	76.24 ± 6.62	328.34
CMC	69.23 ± 5.52	57.54	68.41 ± 5.90	612.27
Votes	96.18 ± 4.32	18.66	97.21 ± 4.51	192.87
WPBC	83.26 ± 3.21	3.11	82.98 ± 3.12	23.45

Table 1 shows that our algorithm’s the training time in the condition of 10-fold cross validation consumed is much less than the RMCLP while their accuracy are in the same level. Generally speaking, our algorithm is faster than both of RMCLP over ten times.

5. Conclusion

In this paper, we have proposed a new algorithm: LP-RMCLP, for binary classification by solving a standard linear programming problem. Experiments have shown that our algorithm is considerably faster, usually over ten times, than RMCLP while the same level of accuracy are kept. Therefore, It is suitable for solving large-scale data sets. Future research includes comprehensive test of the new algorithm and its application.

6. Acknowledgment

This work has been partially supported by grants from National Natural Science Foundation of China(NO.70921061, NO.10601064), the CAS/SAFEA International Partnership Program for Creative Research Teams, Major International-

al(Ragional) Joint Research Project(NO.71110107026), the President Fund of GUCAS.

References

- [1] D. Olson, Y. Shi, Introduction to business data mining, Irwin-McGraw-Hill Series: Operations and Decision Sciences, McGraw-Hill, 2006.
- [2] Y. Shi, Y. Tian, X. Chen, P. Zhang, Regularized multiple criteria linear programs for classification, *Science in China Series F: Information Sciences* 52 (10) (2009) 1812–1820.
- [3] G. Kou, Y. Shi, S. Wang, Multiple criteria decision making and decision support systems - guest editor's introduction, *Decision Support Systems* 51 (2011) 247–249.
- [4] J. Zhang, Y. Shi, P. Zhang, Several multi-criteria programming methods for classification, *Computers and Operations Research* 36 (2009) 823–836.
- [5] Y. Shi, H. Lee, A binary integer linear program with multi-criteria and multi-constraint levels, *Computers & Operations Research* 24 (3) (1997) 259–273.
- [6] J. Li, L. Wei, G. Li, W. Xu, An evolution-strategy based multiple kernels multi-criteria programming approach: The case of credit decision making, *Decision Support System* 51(2):292-298.
- [7] G. Kou, Y. Peng, Y. Shi, M. Wise, W. Xu, Discovering Credit Cardholders' Behavior by Multiple Criteria Linear Programming., *Annals of Operations Research* 135 (1) (2005) 261–274.
- [8] Y. Shi, W. Wise, M. Lou, Multiple criteria decision making in credit card portfolio management, *Multiple Criteria Decision Making in New Millennium*, 2001, pp. 427–436.
- [9] Y. Shi, Y. Peng, W. Xu, X. Tang, Data Mining via Multiple Criteria Linear Programming: Applications in Credit Card Portfolio Management., *International Journal of Information Technology and Decision Making* 1 (1) (2002) 131–152.
- [10] J. Zhang, W. Zhuang, N. Yan, Classification of hiv-1 mediated neuronal dendritic and synaptic damage using multiple criteria linear programming, *Neuroinformatics* 2 (2004) 303–326.
- [11] G. Kou, Y. Peng, Y. Shi, Z. Chen, X. Chen, A multiple-criteria quadratic programming approach to network intrusion detection, in: *CAS-DMKM*, 2004, pp. 145–153.
- [12] W. Kwak, Y. Shi, S. W. Eldridge, G. Kou, Bankruptcy prediction for japanese firms: using multiple criteria linear programming data mining approach., *International Journal of Business Intelligence and Data Mining* 1 (4) (2006) 401–416.
- [13] J. He, Y. Zhang, Y. Shi, G. Huang, Domain-driven classification based on multiple criteria and multiple constraint-level programming for intelligent credit scoring., *IEEE Transactions on Knowledge and Data Engineering* 22 (6) (2010) 826–838.